

Development of Simulation Methods for Membrane Protein Structure Predictions and Replica-Exchange Methods

Ryo URANO^{1,*} and Yuko OKAMOTO^{1,2,3,4}

¹*Department of Physics, Graduate School of Science,
Nagoya University, Nagoya, Aichi 464-8602, Japan*

²*Structural Biology Research Center, Graduate School of Science,
Nagoya University, Nagoya, Aichi 464-8602, Japan*

³*Center for Computational Science, Graduate School of Engineering,
Nagoya University, Nagoya, Aichi 464-8603, Japan*

⁴*Information Technology Center, Nagoya University, Nagoya, Aichi 464-8601, Japan*

Abstract

We developed several simulation methods from two approaches. At first, we extended the previous method for membrane protein structure prediction to treat the flexibility of transmembrane backbone structures, which is often related to functions of proteins. With the new method, we reproduced the native structure of bacteriorhodopsin (BR) which includes distorted transmembrane structures. This method enables us to reduce the system size and computational resource. Secondly, we developed two methods in replica-exchange method (REM). One method is the deterministic replica-exchange method (DETREM), which introduces internal states evolved by an ordinary differential equation which controls replica exchange without pseudo random numbers. This method can resolve some problems of less efficiency in parallel computing caused by pseudo random numbers. The other is the designed walk replica-exchange method (DEWREM), which determines trajectory of replicas in temperature space without a random walk. We applied these new methods and reproduced the results of conventional REM in 2-dimensional Ising system.

1 Introduction

Computers have been developing for decades. However, system of interest has also been larger than before. Thus, efficient algorithm for computer simulation is always important to perform the simulation of current interest. For the purpose, there are two fundamental approaches: decrease of the number of molecule in system and enhancement of sampling. A former example is to adopt an implicit solvent model for solvent in a biomolecule system. In good models, this method can achieve the purpose of simulations although the implicit model abolishes atomistic resolution of solvent molecule

by a replacement of solvent molecule to a mean field function. A latter example is to employ replica-exchange method which enhances sampling of conformations. This method typically accelerates the crossing of free energy barriers using temperature change of replicas.

We developed new methods in each approach, and we show the results of simulations. At first, we proposed a membrane protein structure prediction method with flexible treatment of transmembrane backbone structures and the corresponding extension of the previous implicit membrane model. Using the method, we performed the structure prediction of bacteriorhodopsin and reproduced the native structure. We next proposed two new replica-exchange methods to increase efficiency of REM. One is the method that performs replica exchange with a differential equation without pseudo random numbers. The other is the method that specifies the order of replica exchange and, thus, the trajectory in temperature space among replicas. In 2-dimensional Ising model, we compared results of these new methods and their combination to results of the conventional Metropolis replica-exchange method.

This paper is organized as follows. In the first part, we will introduce the methods and results of membrane structure prediction for bacteriorhodopsin. In the second part, we will show the methods and results of new replica-exchange method in 2D-Ising model. Finally, we will give summary and future prospect of this paper.

2 Methods

We added the following four elementary harmonic constraints as a simple implicit membrane model to the original potential energy function in order to mimic the restrained membrane environment. The constraint energy function is given by

$$E_{\text{constr}} = E_{c1} + E_{c2} + E_{c3} + E_{c4}, \quad (1)$$

where each term is defined as follows:

$$E_{c1} = \sum_{i=1}^{N_H-1} k_1 \theta(r_{i,i+1} - d_{i,i+1}) [r_{i,i+1} - d_{i,i+1}]^2, \quad (2)$$

* Present address: Chemistry Department, Boston University, 590 Commonwealth Avenue, Boston, MA 02215 USA

$$E_{c2} = \sum_{i=1}^{N_H} \left\{ k_2 \theta(|z_i^L - z_{0,i}^L| - d^L) [|z_i^L - z_{0,i}^L| - d^L]^2 + k_2 \theta(|z_i^U - z_{0,i}^U| - d^U) [|z_i^U - z_{0,i}^U| - d^U]^2 \right\}, (3)$$

$$E_{c3} = \sum_{C_\alpha} k_3 \theta(r_{C_\alpha} - d_{C_\alpha}) [r_{C_\alpha} - d_{C_\alpha}]^2, (4)$$

$$E_{c4} = \sum_{j=1}^{N_{BD}} k_4 \theta(|\phi_j - \phi_0| - \alpha_j^\phi) [|\phi_j - \phi_0| - \alpha_j^\phi]^2 + \sum_{j=1}^{N_{BD}} k_5 \theta(|\psi_j - \psi_0| - \alpha_j^\psi) [|\psi_j - \psi_0| - \alpha_j^\psi]^2. (5)$$

E_{c1} is the energy that constrains pairs of adjacent helices along the amino-acid chain not to be apart from each other too much (loop constraints), where $r_{i,i+1}$ is the distance between the C atom of the C-terminus of the i -th helix and the C^α atom of the N-terminus of the $(i+1)$ -th helix, and k_1 and $d_{i,i+1}$ are the force constant and the central value constant of the harmonic constraints, respectively. Each $d_{i,i+1}$ is proportional to the loop length connected between helices. $\theta(x)$ is the step function, which has 1 when x is larger than or equal to 0, otherwise zero. N_H is the total number of transmembrane helices in the protein.

E_{c2} is the energy that constrains helix N-terminus and C-terminus to be located near membrane boundary planes. Here, the z -axis is defined to be the direction perpendicular to the membrane boundary planes. k_2 is the force constant of the harmonic constraints. $z_{0,i}^L$ and $z_{0,i}^U$ are the z -coordinate values of the C^α atom of the N-terminus or C-terminus of the i -th helix near the fixed lower membrane boundary and the upper membrane boundary, respectively. $z_{0,i}^L$ and $z_{0,i}^U$ are the fixed lower boundary z -coordinate value and the upper boundary z -coordinate value of the membrane planes, respectively, and here they depend on each helix atoms due to the known data from OPM[1] although constant membrane plane region is also possible like a previous research condition. d^L and d^U are the corresponding central value constants of the harmonic constraints. This term has a non-zero value only when the C^α atoms of the N-terminus or C-terminus of the i -th helix are apart more than d_i^L (or d_i^U). This constraint energy was introduced so that the helix ends are not too much apart from the membrane boundary planes.

E_{c3} is the energy that constrains all C^α atoms within the sphere (centered at the origin) of radius d_{C^α} . r_{C^α} is the distance of C^α atoms from the origin, and k_3 and d_{C^α} are the force constant and the central value constant of the harmonic constraints, respectively.

E_{c4} is the energy that constrains dihedral angles of main chain within bending or kinked helix structures from ideal helix structures preventing them from forming random-coil structures. ϕ_j and ψ_j are the backbone dihedral angle of the j -th residue. ϕ_0 and ψ_0 are the reference value of the harmonic constraint to keep the helix structures without forming random coil structure, and $\alpha_j^\phi, \alpha_j^\psi$ are the ranges of the harmonic constraints. N_{BD} is the total number of (ϕ, ψ) angles in helix backbones.

We set $k_1 = 5.0$, $d_{i,i+1} = (46, 53, 34, 19, 95, 30)$ where $i=1, 2, \dots, 6$, $k_2 = 5.0$, $z_{0,i}^L = (-14, -16, -20, -15, -19, -24, -18)$ where $i=1, 2, \dots, 7$, $z_{0,i}^U = (12, 14, 15, 15, 14, 11, 12)$ where $i=1, 2, \dots, 7$, $d^U = d^L = 2.0$, $k_3 = 0.5$, $d_{C^\alpha} = 80$, $k_4 = 30.0$, $k_5 = 30.0$, $\phi_0 = -62$, $\psi_0 = -40$, $\alpha_j^\phi = 16$, and $\alpha_j^\psi = 13$. Only the transmembrane helices were used in our simulations, and loop regions of the membrane proteins as well as lipid and water molecules were neglected.

The membrane environment for this protein for the membrane thickness and the region of transmembrane region of the helices was taken from Orientation of Proteins in Membrane (OPM) [1]. The amino-acid sequences of the helices are EWIWLALGTALMGLGLTYFLVKG, KFYAITTLVPAIAFTMYLSMLL, IYWARYADWLFITPLLLLDLALL, QGTILALVGADGIMIGTGLVGL, RFVWVAISTAAMLYLYVLFVGF, TFKVLRNVTVVLWSAYPVVWLVIGSE, and LNIETLLFMVLDVSAKVGFGLLILL. The N-terminus and the C-terminus of each helix were blocked with the acetyl group and the N-methyl group, respectively. The initial structure for each helix was an ideal helix structure and they were placed in the membrane region randomly. We then perform REM simulations of these transmembrane helices. The MC program is based on CHARMM macromolecular mechanics program[2, 3], and replica-exchange Monte Carlo method was implemented in it.

Replica-exchange method is explained in next Methods2 section. We here give the simulation conditions of REM. We used 40 replicas and the following temperatures: 400, 415, 435, 455, 485, 518, 552, 589, 629, 671, 716, 764, 815, 870, 928, 990, 1056, 1127, 1202, 1283, 1369, 1460, 1558, 1662, 1774, 1892, 2019, 2154, 2298, 2452, 2616, 2791, 2978, 3177, 3390, 3616, 3808, 4050, 4250, and 4500 K. We used rather high temperature values compared to experimental conditions. This is because our implicit membrane model guarantees the helix stability and enhances conformational sampling. Replica exchange was attempted at every 50 MC steps. We performed four independent simulations in total of 1,055,950,000 MC steps.

We used the CHARMM19 parameter set (polar hydrogen model) for the potential energy of the system[4, 5]. No cutoff was introduced to the non-bonded terms. Each helix structure was first minimized subjected to harmonic restraint on all the heavy atoms. In order to prepare random initial conformations, we first performed regular

constant MC simulations of all the replicas for 3,000,000 MC steps. We then performed equilibrium MC simulation for 3,000,000 MC steps at the above 40 temperatures, and the last conformation for each replica was the initial structure for the REM simulations. We repeated this process four times for four independent REM simulations. In those simulations, dielectric constant was set to $\epsilon = 1.0$ as in the previous works[6–10]. In MC move, we updated conformations with a rigid translation and rotation of each α -helix, a rotation of torsion angles of backbones by directional manipulation and concerted rotation[11–13], and torsion rotations of side-chains. There are $2N_H + N_{SD} + N_{BD} + N_{CR}$ kinds of MC moves, where N_{SD} is the total number of dihedral angles in the side-chains of N_H helices and N_{CR} is the total number of the combination of seven successive backbone torsion angles by the concerted rotation in the helix backbone. One MC step in this article is defined to be an update of one of these degrees of freedom, which is accepted or rejected according to the Metropolis criterion.

We analyzed the simulation data by the principal component analysis[14–19]. At first, 42,238 conformational data were chosen at each temperature from the REM simulations. The structures were chosen from the trajectories at a fixed interval of 25,000 steps. The structures were superimposed on an arbitrary reference structure, for example, the native structures of PDB code:1PY6. Images were rendered by VMD[20]. The variance-covariance matrix is defined by

$$C_{ij} = \langle (q_i - \langle q_i \rangle)(q_j - \langle q_j \rangle) \rangle, \quad (6)$$

where $\vec{q} = (q_1, q_2, q_3, \dots, q_{3n-1}, q_{3n}) = (x_1, y_1, z_1, \dots, x_n, y_n, z_n)$ and $\langle \vec{q} \rangle = \sum_{k=1}^n \vec{q}(k)/n$. x_i, y_i, z_i are Cartesian coordinates of the i -th atom, and n is the total number of atoms.

This calculation was performed by R program package[21], and the clustering was performed by k-means clustering method[22]. The first superposition was done to remove large eigenvalues from the translations and rotations of the system because we want to analyze the internal differences of structures. The eigenvalues were ordered in the decreasing order of magnitude.

3 Results1

We first identified the free energy minimum state in our simulations classified by principal component analysis. Fig. 1 shows the representative structure in each cluster from the highest density region. The root-mean-square-deviation (RMSD) value of each representative structure with respect to the C^α atoms was 3.6 Å, 8.8 Å, 15.8 Å, 15.9 Å, and 16.6 Å for Cluster 1, Cluster 2, Cluster 3, Cluster 4, and Cluster 5, respectively. From these RMSD values, we found that the native-like structure is the second-lowest free energy state (Cluster 1) and that

the global-minimum free energy state (Cluster 2) is the second closest to the native structure. In the structure of Cluster 2, the space where the retinal molecule occupies in the native structure is filled with a helix, and this increases the contact between helices and seems to stabilize this structure more than the native-like structure of Cluster 1 with the empty space for the retinal molecule. Moreover, the result that a helix occupies the retinal space is consistent with previous works[6, 10] which did not include the flexibility of helix structures. However, the previous works were not able to obtain the native-like structure such as Cluster 1. Hence, the extension of including the freedom of helix structure distortion has improved the accuracy of prediction for membrane protein structure determination by simulation. Our results suggest that in the simulations without a retinal molecule the structures can interchange between the structures of Cluster 1 and Cluster 2. After an insertion of a retinal, it then stabilizes the native-like structure as shown in Fig.2. It is important that the association of helices enabled them to make a room for an insertion of a retinal molecule. This is consistent with the experimental results of bacteriorhodopsin, which observed the spontaneous insertion of a retinal molecule by a helix association[23].

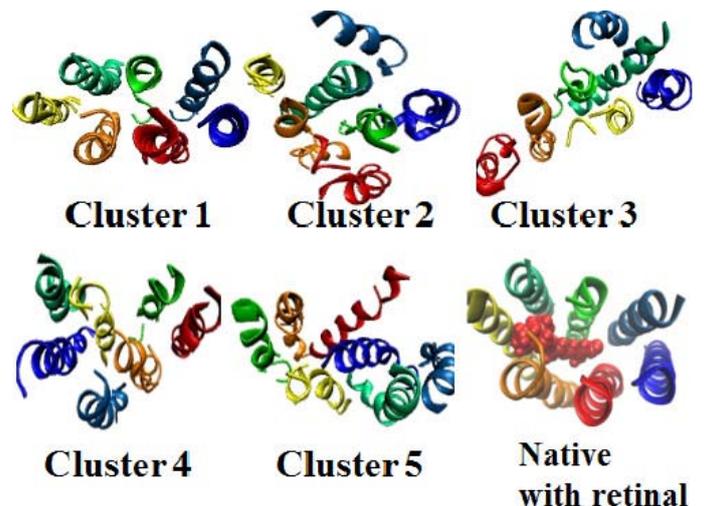


FIG. 1. Typical structures in each cluster selected in the highest density region. The RMSD from the native conformation with respect to all C^α atoms is 3.6 Å, 8.8 Å, 15.8 Å, 15.9 Å, and 16.6 Å for Cluster 1, Cluster 2, Cluster 3, Cluster 4, and Cluster 5, respectively. Helices are colored from the N-terminus to the C-terminus: blue (Helix A), lightblue (Helix B), green (Helix C), deepgreen (Helix D), yellow (Helix E), orange (Helix F), and red (Helix G).

4 Methods2

We now give the details of our new replica-exchange methods. We prepare M non-interacting replicas at M different temperatures. Let the label i ($=1, \dots, M$) stand

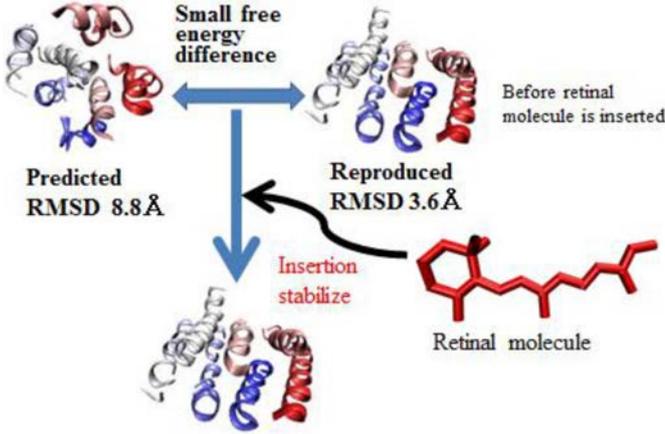


FIG. 2. Hypothesis about the relation between the global-minimum free energy state and the second-minimum. The effect of an insertion of a retinal molecule causes the stabilization of the native-like structure.

for the replica index and label m ($=1, \dots, M$) for the temperature index. We represent the state of the entire system of M replicas by $X = \{x_{m(1)}^{[1]}, \dots, x_{m(M)}^{[M]}\}$, where $x_m^{[i]} = \{q^{[i]}, p^{[i]}\}_m$ are the set of coordinates $q^{[i]}$ and momenta $p^{[i]}$ of particles in replica i (at temperature T_m). The probability weight factor for state X is given by a product of Boltzmann factors:

$$W_{\text{REM}}(X) = \prod_{i=1}^M \exp[-\beta_{m(i)} H(q^{[i]}, p^{[i]})], \quad (7)$$

where $\beta_m (= 1/k_B T_m)$ is the inverse temperature and $H(q, p)$ is the Hamiltonian of the system. We consider exchanging a pair of replicas i and j corresponding to temperatures T_m and T_n , respectively:

$$X = \{\dots, x_m^{[i]}, \dots, x_n^{[j]}, \dots\} \rightarrow X' = \{\dots, x_n^{[i]}, \dots, x_m^{[j]}, \dots\}, \quad (8)$$

where $x_n^{[i]} \equiv \{q^{[i]}, p^{[i]}\}_n$, $x_m^{[j]} \equiv \{q^{[j]}, p^{[j]}\}_m$, and $p^{[j]'} = \sqrt{\frac{T_m}{T_n}} p^{[j]}$, $p^{[i]'} = \sqrt{\frac{T_n}{T_m}} p^{[i]}$ [24].

Here, the transition probability $\omega(X \rightarrow X')$ of Metropolis criterion for replica exchange is given by

$$\begin{aligned} \omega(X \rightarrow X') &= \min\left(1, \frac{W_{\text{REM}}(X')}{W_{\text{REM}}(X)}\right) \\ &= \min(1, \exp(-\Delta)), \end{aligned} \quad (9)$$

where

$$\Delta = \Delta_{m,n} = (\beta_n - \beta_m)(E(q^{[i]}) - E(q^{[j]})). \quad (10)$$

Because each replica visits various temperatures followed by the transition probability of Metropolis algorithm, REM performs a random walk in temperature space.

We now review two REMs, which are based on random walks in temperature space. Without loss of generality, we can assume that M is an even integer and that

$T_1 < T_2 < \dots < T_M$. The conventional REM[24–27] is performed by repeating the following two steps:

1. We perform a conventional MD or MC simulation of replica i ($= 1, \dots, M$) at temperature T_m ($m = 1, \dots, M$) simultaneously and independently for short steps.
2. Pairs of exchange attempts are selected in replica pairs with neighboring temperatures, for example, for the odd pairs $(T_1, T_2), (T_3, T_4), \dots, (T_{M-1}, T_M)$ or even pairs $(T_2, T_3), (T_4, T_5), \dots, (T_{M-2}, T_{M-1})$.

All the replica pairs thus selected are attempted to be exchanged according to the Metropolis transition probability in Eqs. (9) and (10) with $n = m + 1$.

We repeat Steps 1 and 2 until the end of the simulation. The canonical ensemble at any temperature is reconstructed by reweighting techniques[28–30].

We next present the deterministic replica-exchange method (DETREM)[31]. Only Step 2 is different from the conventional REM. At first, we introduce an internal state $y_{m,n}$ as an index of a pair of replicas i and j at temperatures T_m and T_n , and consider the following differential equation:

$$\frac{dy_{m,m+1}}{dt} = \sigma_m \frac{1}{1 + \exp(\Delta_{m,m+1})}, \quad (11)$$

where t is a virtual time, $\Delta_{m,m+1}$ is the same as in Eq. (10) with $n = m + 1$, and the signature σ_m of the pair of (T_m, T_{m+1}) changes to 1 or -1 to control the signature of the change of y_m which monotonically increases or decreases. In Step 2, instead of applying the Metropolis criterion in Eqs. (9) and (10), we solve the differential equation in Eq. (11) for the internal states $y_{m,m+1} \in \{-1, 1\}$ for (T_m, T_{m+1}) , where the total number of internal states is $M-1$ with the following pairs: $(1,2), (2,3), \dots, (M-1,M)$ for the random-walk DETREM and the pairs: $(1,2), (3,4), \dots, (M-1,M)$ and $(2,3), (4,5), \dots, (M-2,M-1)$ for designed-walk REM. The replica exchange is done as follows[31]:

if updated $y_{m,m+1} \geq \pm 1$, then $(T_m, T_{m+1}) \rightarrow (T_{m+1}, T_m)$,
 $y_{m,m+1} \leftarrow y_{m,m+1} \mp 1$, and $\sigma_m \leftarrow \mp 1$.

For the random-walk DETREM, if $y_{m,m+1}$ performs exchanges, $y_{m+1,m+2}$ is not time evolved and $y_{m+2,m+3}$ is evolved to avoid the leap exchange of temperature such as from T_m to T_{m+2} .

Finally, the temperature walk can be implemented to both designed conventional REM and DETREM (and other REMs) as follows. Namely, the designed-walk replica-exchange method (DEWREM)[32] is performed by repeating the following steps.

1. We perform a conventional MD or MC simulation of replica i ($= 1, \dots, M$) at temperature T_m ($m =$

1, \dots , M) simultaneously and independently for short steps.

2. Replica exchange is attempted for all the odd pairs (T_1, T_2) , (T_3, T_4) , \dots , (T_{M-1}, T_M) .

3. Repeat Steps 1 and 2 until all odd pairs perform replica exchange exactly once. Namely, once a pair is exchanged, the exchanged pair stops exchange attempts and keep performing the simulation in Step 1 with the new temperatures. Replica exchange attempt in Step 2 is repeated until all the other odd pairs finish exchanges.

4-6. Repeat Steps 1-3 where the odd pairs in Steps 2 and 3 are now replaced by the even pairs (T_2, T_3) , (T_4, T_5) , \dots , (T_{M-2}, T_{M-1}) .

7. The cycle of Steps 1 to 6 is repeated until the number of cycles is M , which is equal to the tunneling count and all replicas have the initial temperatures.

8. Begin the above cycle of Steps 1-7 with Steps 1 to 3 and Steps 4 to 6 interchanged. These eight steps are repeated until the end of the simulation.

The schematic picture of this procedure is shown in Fig. 3. We remark that Step 8, namely, reversing the cycle of Steps 1-3 and 4-6, is necessary for the detailed balance condition, because the entering states are the same as leaving states. For example, the state $(x_1^{[1]}, x_2^{[2]}, x_3^{[3]}, x_4^{[4]}, x_5^{[5]}, x_6^{[6]})$ is reached from only two states $(x_2^{[1]}, x_1^{[2]}, x_4^{[3]}, x_3^{[4]}, x_6^{[5]}, x_5^{[6]})$, $(x_1^{[1]}, x_3^{[2]}, x_2^{[3]}, x_5^{[4]}, x_4^{[5]}, x_6^{[6]})$ only and makes transition to the two states. as shown in Fig. 3, where $x_m^{[i]}$ is the state of replica i at temperature T_m . This exchange procedure satisfies the detailed balance condition for replica and temperature pair because the trial of exchange pair

$$\begin{aligned} & \gamma(i(m) \rightarrow i(m+1)) \times \\ & \omega((x_m^{i(m)}, x_{m+1}^{i(m+1)}) \rightarrow (x_{m+1}^{i(m)}, x_m^{i(m+1)})) \\ & = \gamma(i(m+1) \rightarrow i(m)) \times \\ & \omega((x_{m+1}^{i(m)}, x_m^{i(m+1)}) \rightarrow (x_m^{i(m)}, x_{m+1}^{i(m+1)})) \end{aligned} \quad (12)$$

is equal in the route as is shown in Fig. 3, where $\gamma(i(m) \rightarrow i(m+1))$ is the selected probability of the exchange attempt.

This sequential exchange achieves one tunneling count when M cycles for each replica are finished. In theory, the estimated ratio of tunneling count between the odd-even sequential exchange and the conventional random walk is given by

$$\frac{TC_{\text{sequential}}}{TC_{\text{random walk}}} = \frac{N_{\text{trial}} \times P_{\text{correction}}^{\text{DEW}}}{\sqrt{N_{\text{trial}}} \times P_{\text{correction}}^{\text{RW}}} \propto \sqrt{N_{\text{trial}}}, \quad (13)$$

where N_{trial} is the number of exchange attempts, $P_{\text{correction}}^{\text{DEW}}$ is the correction for waiting for all the replica

exchanges in Steps 3 and 6, and $P_{\text{correction}}^{\text{RW}}$ is the correction for the deviation of random-walk probability from the value 1/2.

In order to test the effectiveness of the present methods, we performed simulations with conventional REM, DETREM and DEWREM for a 2-dimensional Ising model. The lattice size in a square lattice was 128 (hence, the number of spins was $N = 128^2 = 16384$). We have performed conventional random-walk simulation and designed-walk simulation of both Metropolis REM and DETREM. We have also performed a mixed random-walk and designed-walk simulation of DETREM, where we repeated the two walks alternately. The total number of replicas M was 40 and the temperatures were 1.50, 1.55, 1.60, 1.65, 1.70, 1.75, 1.80, 1.85, 1.90, 1.94, 1.98, 2.01, 2.04, 2.07, 2.10, 2.13, 2.16, 2.19, 2.22, 2.25, 2.28, 2.31, 2.34, 2.358, 2.368, 2.38, 2.40, 2.42, 2.44, 2.47, 2.51, 2.57, 2.63, 2.69, 2.75, 2.82, 2.90, 3.00, 3.10, and 3.15. Boltzmann constant k_B and coupling constant J were set to 1. Thus, $\beta = 1/k_B T = 1/T = \beta^*$, and the potential energy is given by $E(\mathbf{s}) = -\sum_{\langle i,j \rangle} s_i s_j$, where $s_i = \pm 1$, and the summation is taken over all the nearest-neighbor pairs in the square lattice.

For the conventional random-walk REM and DETREM, replica-exchange attempt was made every 1 MC step. One MC step here consists of one Metropolis update of spins. The total number of MC steps for all the simulations was 100,000,000. To integrate Eq. (11), we used the fourth-order Runge-Kutta method with virtual time step $dt = 1$. For DEWREM, replica-exchange attempt was made every 20, 50, 100, 150, 200 MC steps in the conventional REM simulations and every 20, 50, 100, 150, 200, and 250 MC steps in the DETREM simulations (see Table I). The mixed-walk simulation was performed in which after $4M (= 160)$ even-odd or odd-even cycles of designed-walk simulations (replica-exchange attempt was made at every 20 MC steps) were performed, 200,000 MC steps (which roughly corresponds to $2M$ cycles) of random-walk simulations (replica-exchange attempt was made at every 1 MC step) were performed, and then this procedure was repeated. For reweighting analyses ([28-30, 33]), the total of 10,000 spin state data were taken with a fixed interval of 1,000 MC steps at each temperature from the REM simulations.

5 Results2

Table I lists the mean tunneling counts per replica for each method, which is the number of times where the replicas visit from the lowest temperature through the highest temperature and back to the lowest during the simulation. The mean tunneling counts per replica of the designed-walk simulations at every 10 MC attempts were about twice larger. These large tunneling counts imply that in designed-walk method all replicas traversed

more efficiently in temperature space, and our design to maximize the tunneling counts for all replicas without random walks was successful. For the mixed-walk simulation, the maximum tunneling count was about twice larger than that of random-walk DETREM. The mean tunneling count was almost the same as that of designed-walk DETREM.

We next examine physical quantities obtained from the designed-walk simulations with various replica-exchange attempt frequencies and mixed walk simulations and compare them to those from the conventional random-walk simulations. Fig. 4(a) and Fig. 4(b) show the specific heat C as a function of T during the conventional REM simulations and the DETREM simulations, respectively. They were obtained by the reweighting techniques[28–30, 33]. This shows that designed-simulation with shorter replica-exchange interval such as every 10 and 20 MC steps underestimated the heat capacity near the critical temperature although the transition point is sufficiently similar to the exact critical temperature at $T_c = 2.269$. As the intervals of replica-exchange attempts are longer, the accuracy of heat capacity is higher. Moreover, the combination of the random-walk and designed walk also increased the accuracy. This suggests that the designed-walk replica-exchange attempts caused correlation between replicas. The correlation seems to be very strong near the critical temperature. As a result, the heat capacity is underestimated slightly. Fig. 5(a) and Fig. 5(b) show susceptibility χ as a function of temperature obtained from the random-walk and designed-walk simulations of Metropolis REM and DETREM, for DETREM including the mixed-walk simulation. This figure shows that by extending intervals of replica-exchange attempts DEWREM simulation can reproduce the results of random-walk REM in both conventional REM and DETREM. However, we observe slower relaxation of susceptibility in different replica exchange intervals to the conventional results than that of heat capacity. Moreover, these physical quantities show that repeating a random walk and designed walk simulation in mixed-walk simulation is an efficient way to increase the accuracy of results and the number of tunneling counts at the same time.

However, these physical quantities show that mixed-walk simulation can increase the accuracy of results and the number of tunneling counts and DEWREM simulation is suited for simulations with longer time intervals between replica-exchange attempts.

6 Summary and future prospect

The first part of this article gave the results of protein structure prediction for bacteriorhodopsin with the flexible treatment of transmembrane helix backbone structure and our recently extended implicit membrane model.

We obtained not only the native-like structure but also the associated structure with an empty space for retinal molecule insertion. These structures are also consistent with previous experimental results.

In the next part, we compared the deterministic replica-exchange method and designed walk replica-exchange method to the conventional REM in 2-dimensional Ising model. DETREM exactly reproduced the results of conventional REM including phase transition. On the other hand, DEWREM needs longer interval between replica-exchange trials because of correlation caused by the introduction of designed walk, which may break the Markov process in short interval resulting in the deviation from the exact value near the critical temperature. To avoid this correlation, mixing random walk is a better way to remove the correlation.

Finally, we will give some perspectives for future work. We will first apply our structure prediction method to unknown membrane proteins which obtained the low resolution structure.. Secondly, DETREM can introduce the methods for faster convergence in machine learning because this method has the same mathematical formulation of Boltzmann machine. Thirdly, the way for fast reduction of correlation among replicas in shorter interval is useful in the application of spin systems in DEWREM while another system such as a peptide usually employs longer time interval in REM.

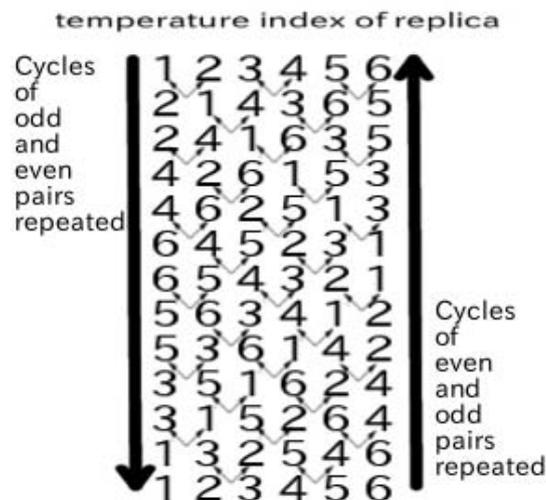


FIG. 3. An schematic picture of time series of temperature indices in DEWREM with 6 replicas. The left cycle begins with the temperature exchange of odd index pairs (T_1, T_2), (T_3, T_4), and (T_5, T_6), then tries with even pairs (T_2, T_3) and (T_4, T_5). The right cycle begins with even pairs and next tries odd pairs. They are the reverse cycles of each other and their combination satisfies the detailed balance condition of replica exchange.

Acknowledgments

TABLE I. The mean number of tunneling counts per replica.

TC	Random walk		Designed walk								Mixed walk
	Met	DETREM	Met				DETREM				DETREM
Interval	1	1	20	50	100	150	20	100	150	200	1 & 20
Mean	173	178	292	197	131	99	231	93	69	55	293
\pm SD	9.5	8.8	56	41	27	21	48	20	15	11	6

TC, Interval, SD, Met stand for tunneling counts, the number of MC steps between replica-exchange attempts, standard deviation, and REM based on Metropolis criterion, respectively. The frequency (1 & 20) of Mixed walk means that it was 1 MC step for random walk REM and 20 MC steps for designed-walk REM.

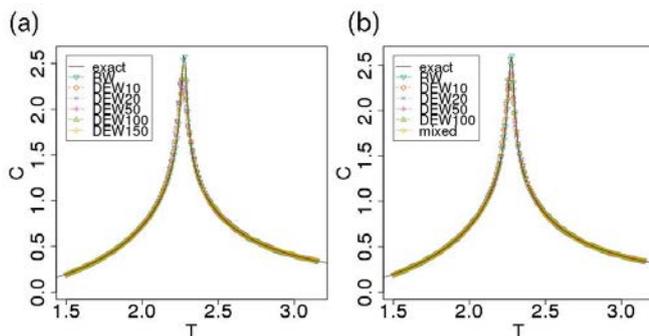


FIG. 4. Specific heat C as a function of T from the (a) REM, (b) DETREM simulations including the mixed-walk simulation. The error bars are smaller than the symbols. In the inset, the labels are as follows. exact: exact solution, RW: random walk, DEW n : DEWREM with the interval of n MC steps, and mixed: mixed walk (see Table 1). The exact results for the lattice size $L = 128$ (black curves) were obtained by Berg's program [34] based on Ref. [35].

Some of the computations were performed on the supercomputers at the Institute for Solid State Physics, University of Tokyo. This work was supported, in part, Grants-in-Aid for Scientific Research (A) (No. 25247071), for Scientific Research on Innovative Areas ("Dynamical Ordering & Integrated Functions"), Program for Leading Graduate Schools "Integrative Graduate Education and Research in Green Natural Sciences", and for the Computational Materials Science Initiative, and for High Performance Computing Infrastructure from the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan.

-
- [1] M. Lomize, A. Lomize, I. Pogozheva, and H. Mosberg, *Bioinformatics* **22**, 623 (2006).
 [2] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, *J. Comput. Chem.* **4**, 187 (1983).

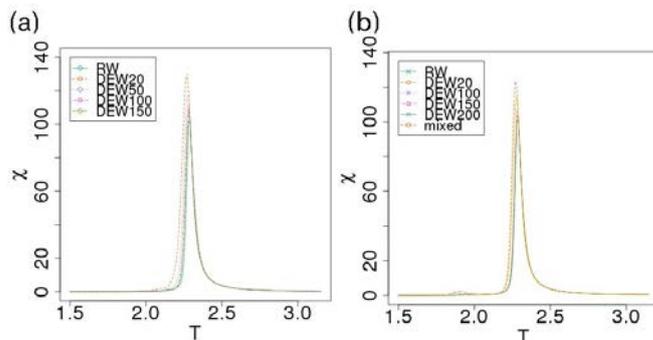


FIG. 5. Susceptibility χ as a function of T from the (a) REM, (b) DETREM simulations including the mixed-walk simulation. The error bars are smaller than the symbols. Also see the caption of Fig. 4.

- [3] J. Hu, A. Ma, and A. R. Dinner, *J. Comput. Chem.* **27**, 203 (2006).
 [4] W. Reiher, *Theoretical studies of hydrogen bonding*, Ph.D. thesis, Harvard University (1985).
 [5] E. Neria, S. Fischer, and M. Karplus, *J. Chem. Phys.* **105**, 1902 (1996).
 [6] H. Kokubo and Y. Okamoto, *Biophys. J.* **96**, 765 (2009).
 [7] H. Kokubo and Y. Okamoto, *Chem. Phys. Lett.* **383**, 397 (2004).
 [8] H. Kokubo and Y. Okamoto, *J. Chem. Phys.* **120**, 10837 (2004).
 [9] H. Kokubo and Y. Okamoto, *J. Phys. Soc. Jpn.* **73**, 2571 (2004).
 [10] H. Kokubo and Y. Okamoto, *Chem. Phys. Lett.* **392**, 168 (2004).
 [11] A. R. Dinner, *J. Comput. Chem.* **21**, 1132 (2000).
 [12] N. Gō and H. A. Scheraga, *Macromolecules* **3**, 178 (1970).
 [13] L. Dodd, T. Boone, and D. Theodorou, *Mol. Phys.* **78**, 961 (1993).
 [14] M. Teeter and D. Case, *J. Phys. Chem.* **94**, 8091 (1990).
 [15] A. Kitao, F. Hirata, and N. Gō, *Chem. Phys.* **158**, 447 (1991).
 [16] A. Garcia, *Phys. Rev. Lett.* **68**, 2696 (1992).
 [17] R. Abagyan and P. Argos, *J. Mol. Biol.* **225**, 519 (1992).
 [18] A. Amadei, A. Linssen, and H. Berendsen, *Proteins* **17**, 412 (1993).
 [19] A. Kitao and N. Gō, *Curr. Opin. Struct. Biol.* **9**, 164 (1999).
 [20] W. Humphrey, A. Dalke, and K. Schulten, *J. Mol. Graphics* **14**, 33 (1996).
 [21] R. Ihaka and R. Gentleman, *J. Comput. Graph. Stat.* **5**, 299 (1996).
 [22] J. MacQueen, in *Fifth Berkeley Symposium on Mathematical Statistics and Probability*, eds Le Cam LM, Neyman J (Univ of California Press, Berkeley, CA) (1967) pp. 281–297.
 [23] J. L. Popot, S. E. Gerchman, and D. M. Engelman, *J. Mol. Biol.* **198**, 655 (1987).
 [24] Y. Sugita and Y. Okamoto, *Chem. Phys. Lett.* **314**, 141 (1999).
 [25] K. Hukushima and K. Nemoto, *J. Phys. Soc. Jpn.* **65**, 1604 (1996).
 [26] R. H. Swendsen and J.-S. Wang, *Phys. Rev. Lett.* **57**, 2607 (1986).

- [27] C. J. Geyer, *Comput. Sci. Stat.: Proc. 23rd Symp. Interface*, Interface Foundation, Fairfax Station, VA , 156 (1991).
- [28] A. M. Ferrenberg and R. H. Swendsen, *Phys. Rev. Lett.* **63**, 1195 (1989).
- [29] S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman, *J. Comput. Chem.* **13**, 1011 (1992).
- [30] A. Mitsutake, Y. Sugita, and Y. Okamoto, *J. Chem. Phys.* **118**, 6664 (2003).
- [31] R. Urano and Y. Okamoto, e-print: arXiv:1412.6959 (2014).
- [32] R. Urano and Y. Okamoto, e-print: arXiv:1501.00772 (2015).
- [33] M. R. Shirts and J. D. Chodera, *J. Chem. Phys.* **129**, 124105 (2008).
- [34] B. A. Berg, *Markov Chain Monte Carlo Simulations and Their Statistical Analysis* (World Scientific, Singapore, 2004).
- [35] A. E. Ferdinand and M. E. Fisher, *Phys. Rev.* **185**, 832 (1969).